Mol. Biol. (1962) 5, 139–158

The Enzymic Synthesis of Amino Acyl Derivatives of Ribonucleic Acid

V. Nucleotide Sequences Adjacent to the . .pCpCpA End Groups

ULF LAGERKVIST† AND PAUL BERG

Department of Biochemistry, Stanford University School of Medicine, Palo Alto, California, U.S.A.

(*Received 25 April 1962*)

The composition and sequence of nucleotides occurring next to the common terminal trinucleotide sequence, . .pCpCpA, of amino acid-acceptor ribonucleic acid chains has been analysed. The normally existing . .pCpCpA end-groups were enzymatically removed, replaced by . .^{32}pC^{32}pC and the labeled ribonucleic acid was degraded with alkali, pancreatic and T_1 ribonuclease. The analysis of the labeled fragments produced in each digestion demonstrates that, aside from the common . .pCpCpA terminal sequence, there is considerable heterogeneity in the nucleotide sequences at the amino acid-acceptor end of the chains. Eleven different sequences, ranging in length from 4 to 9 nucleotides and accounting for approximately 80% of the ribonucleic acid chains in the population, have been identified.

1. Introduction

Among the cellular RNAs only one class acts as an amino acid acceptor (see Hoagland, 1960, and Berg, 1961, for recent reviews). The amino acid-acceptor RNA fraction is a mixture of chains each specific for a single amino acid. In each case the amino acid is linked to its appropriate RNA chain by an ester bond to an hydroxyl group of the terminal nucleotide (adenyl) residue (Zachau, Acs & Lipmann, 1958; Preiss, Berg, Ofengand, Bergmann & Dieckmann, 1959; Hecht, Stephenson & Zamecnik, 1959). Moreover, separate enzymes, amino acyl RNA synthetases (Berg, Bergmann, Ofengand & Dieckmann, 1961), catalyse the formation of each amino acyl RNA derivative (Berg & Ofengand, 1958; Allen, Glassman & Schweet, 1960; Lipmann, Hülsmann, Hartmann, Boman & Acs, 1959). At present little is known of the structural parameters of the RNA which distinguish one amino acid-specific RNA chain from another. A possible clue is the recent finding that RNA preparations enriched for alanine-, valine-, and tyrosine-acceptor activity have different base compositions (Holley, Apgar, Merrill & Zubkoff, 1961).

Inasmuch as each amino acid-acceptor RNA chain contains an identical trinucleotide sequence, adenylyl 5′→3′ cytidylyl 5′→3′ cytidylyl 5′→. . . . at the acceptor end (Hecht, Zamecnik, Stephenson & Scott, 1958; Canellakis & Herbert, 1960; Preiss, Dieckmann & Berg, 1961; Furth, Hurwitz, Krug & Alexander, 1961) and a guanosine-5′-phosphate residue at the other end (Singer & Cantoni, 1960; Zillig, Schachtschabel & Krone, 1960), these portions of the chain may be excluded as the

† Rockefeller Foundation Fellow; permanent address: University of Gothenburg, Gothenburg, Sweden.

sole origins of the specificity. One possibility is that the specificity resides in nucleotide sequences adjacent to or near the pCpCpA† segment. To test this hypothesis we have examined two questions. First, is there sufficient heterogeneity in the nucleotide sequences adjacent to the pCpCpA end groups to specify the twenty or more amino acid-specific chains and second, are different amino acid-acceptor RNA chains distinguishable by the nucleotide sequences in this region? Our results establish that, aside from the common pCpCpA end groups, marked differences in the 3′-hydroxy terminal nucleotide sequences do occur. Eleven different sequences, accounting for approximately 80% of the acceptor RNA chains and ranging in length from 4 to 9 nucleotides, have been identified. The next paper (Berg, Lagerkvist & Dieckmann, 1962) shows that one of these, ..GpCp(UpC)pApCpCpA, arises from RNA chains which accept isoleucine and two others, ..GpCpApCpCpA and ..GpUpApCpCpA, occur in RNA chains which accept leucine.

2. Basic Experimental Approach

The method relies on selective labeling of the RNA chains at the amino acid-acceptor end and analysis of the labeled fragments produced by specific degradation procedures.

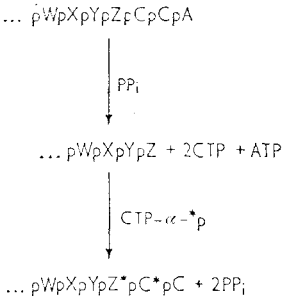

FIG. 1. Pyrophosphorolysis and subsequent labeling of amino acid-acceptor RNA.

1. RNA was enzymically pyrophosphorolysed to remove the terminal three nucleotides from each RNA chain (Preiss *et al.*, 1961) and then two CM*P residues were added to form RNA..*pC*pC (Fig. 1).

2. Digestions of the labeled RNA with alkali, pancreatic RNase and takadiastase RNase (T_1 RNase) produce a mixture of mono-, di-, and oligonucleotide fragments; in each case, the fragments originating from the amino acid-acceptor end of each chain are labeled with *P and therefore are distinguishable from fragments arising from internal regions of the RNA chains (Fig. 2).

† The following abbreviations have been used throughout: A for adenosine, G for guanosine, C for cytidine, U for uridine, ψU for 5-ribosyl uracil (pseudo-uridine), and rT for 5-methyluridine. For shorthand purposes in designating nucleotide sequences or the products derived from the degradation of oligonucleotides, the following designations have been used (Markham & Smith, 1952): Ap, Gp, Cp, Up, rTp and ψUp for the 2′- or 3′-nucleoside monophosphates (or a mixture of both) and pA, pG, pC and pU for the 5′-nucleoside monophosphates. PP_i has been used to denote inorganic pyrophosphate. Where a phosphate residue is labeled with ^{32}P it has been designated with a *, e.g. *P, *pA or A*p; where the purine or pyrimidine is labeled with ^{14}C it has been designated with a •, e.g. •Ap or p•A; when a particular nucleotide contains both ^{32}P and ^{14}C the following symbol has been used: •A*p. The intact structure of amino acid-acceptor RNA has been written RNA..pCpCpA and the enzyme which reversibly pyrophosphorolyses the terminal trinucleotide as RNA..(pCpCpA) pyrophosphorylase. PDE and PME have been used to designate phosphodiesterase and phosphomonoesterase, respectively.

5. The fragments produced in each digestion were isolated and characterized and the corresponding terminal nucleotide sequences existing in the intact acceptor RNA chains were deduced.

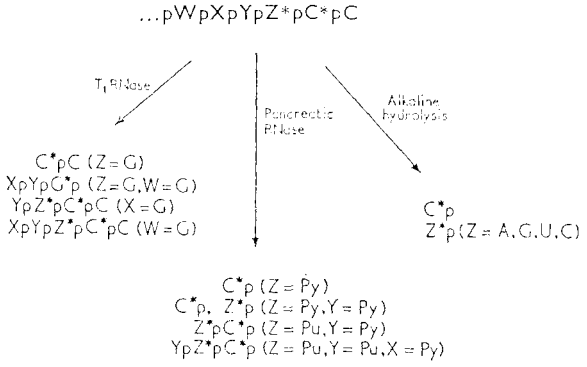

FIG. 2. Schematic representation of the types of fragments produced by alkaline, pancreatic RNase and T_1 RNase digestion of terminally labeled amino acid-acceptor RNA.

3. Detailed Experimental Procedure

(a) *Materials*

[2-^{14}C]uracil and [2-8-^{14}C]adenine were purchased from the California Corporation for Biochemical Research.

CTP labeled with *P in the ester phosphate (4×10^7 cts/min/μmole) was prepared enzymatically (Preiss *et al.*, 1961) from CM*P which was synthesized chemically according to Hurwitz (1959). Uniformly labeled [^{14}C]CTP ($9{\cdot}8 \times 10^5$ cts/min/μmole) was also prepared enzymatically from the corresponding uniformly labeled [^{14}C]CMP isolated from the RNA of *Chromatium* grown on $^{14}CO_2$ as sole carbon source (Ofengand, 1959). [8-^{14}C]ATP ($6{\cdot}5 \times 10^5$ cts/min/μmole) was purchased from Schwarz Laboratories. A mixture of the 5′-, 3′- and 2′-phosphates of [2-^{14}C]ribothymidine (rT) was made by chemical phosphorylation. The rT was prepared enzymatically from [2-^{14}C]thymine (purchased from the Radiochemical Centre, Amersham, England) and unlabeled uridine using a nucleoside phosphorylase preparation from *E. coli* (Lagerkvist, unpublished experiments). 5-Ribosyl uracil 2′-5′-monophosphate (ψUp) was a gift and 5-methyl deoxyuridine 3′-monophosphate (dTp) was isolated by paper electrophoresis from a spleen PDE digest of calf thymus DNA (Josse, Kaiser & Kornberg, 1961).

RNA..(pCpCpA) pyrophosphorylase was prepared from *E. coli* as previously described (Preiss *et al.*, 1961); the activity (units) is expressed on the basis of CMP incorporation. The same preparations were used for pyrophosphorolysis and for the incorporation of CMP and AMP. Pancreatic RNase was the crystalline preparation obtained from Worthington Biochemical Corporation and T_1 RNase was isolated from takadiastase (Takahashi, 1961) and contained 1530 units/mg of protein. Snake venom PDE free of PME activity was isolated from the venom of *Crotalus adamanteus* according to Koerner & Sinsheimer (1957) and contained 1600 units/ml. (1 unit equals 1 μmole of phosphodiester bonds split per hour at 37°C in a DNase-limit digest of DNA (Lehman, Bessman, Simms & Kornberg, 1958)). *E. coli* PME (Garen & Levinthal, 1960) was purchased from Worthington Biochemical Corporation and contained 2200 units/ml. (1 unit equals 1 μmole P_i released from AMP/hr at 37°C). RNase from *Bacillus subtilis*, reported to be specific for the cleavage of purine ribonucleoside phosphodiester linkages (Nishimura, 1960), was a gift.

Unlabeled amino acid-acceptor RNA was isolated from *E. coli* by the method of Ofengand, Dieckmann & Berg (1961). To facilitate the identification of the oligonucleotide fragments produced by T_1 RNase, the starting amino acid-acceptor RNA was labeled with [^{14}C]purines and pyrimidines throughout the chains. The [^{14}C]RNA was obtained

from two *E. coli* mutants. One (W3362) requiring a pyrimidine and the other (W3687) a purine were grown separately in a glucose–salts medium supplemented with [2-^{14}C]uracil or [2-^{14}C]adenine, respectively. After maximal growth was reached (>95% of the labeled purine or pyrimidine was incorporated into the cells), the cultures were pooled, the cells were harvested, mixed with unlabeled cells of strains W3362 and W3687 and then the acceptor RNA was isolated as described by Ofengand *et al.* (1961). The specific ^{14}C activity of the isolated purine and pyrimidine nucleotides produced by alkaline hydrolysis was 2·9 and $3·9 \times 10^8$ cts/min/μmole, respectively. RNA concentrations are expressed in terms of its mononucleotide content (Ofengand *et al.*, 1961).

(b) *Methods*

(i) *Preparation of RNA..*pC*pC.*

(a) Pyrophosphorolysis of RNA..pCpCpA—the standard reaction mixture contained per ml.: 50 μmoles of tris buffer, pH 7·5, 6 μmoles of $MgCl_2$, 10 μmoles of 2-mercaptoethanol, 5 μmoles of KF, 5 μmoles of sodium pyrophosphate, pH 7·5, 0·2 μmole of RNA and 2 units of RNA..(pCpCpA) pyrophosphorylase. The mixture (usually containing 10 to 30 μmoles of RNA) was incubated at 37°C for 2 hr and then the RNA was precipitated by the addition of NaCl (to 1·5 M) and 2 vol. of ice-cold ethanol. After 1 hr at 3°C the precipitate was collected by centrifugation, washed twice with a salt–ethanol mixture (0·5 M-NaCl–67% ethanol) and then extracted twice with 3 ml. portions of cold water. The combined extracts were dialysed for 36 hr against 3 changes of 6 l. of 0·2 M-NaCl and then for 36 hr against 2 to 3 changes of water. The recovery of RNA was approximately 85%.

Pyrophosphorolysis of the terminal three nucleotides, which was performed twice, was judged to be complete because: (1) the amount of CMP added back to the pyrophosphorolysed RNA preparations was equivalent to 2 residues per average chain length of 89 nucleotides (range of 4 preparations = 83 to 95), and (2) the ratio of CMP to AMP incorporation into the pyrophosphorolysed RNA preparation was 1·95 to 2·15.

(b) Labeling of pyrophosphorolysed RNA with CM*P— the standard reaction mixture contained per ml.: 40 μmoles of either tris buffer pH 7·5 or potassium phosphate buffer, pH 7·6, 10 μmoles of $MgCl_2$, 10 μmoles of 2-mercaptoethanol, 0·4 μmole of pyrophosphorolysed RNA, 0·04 μmole of CT*P (4×10^7 cts/min/μmole) and about 3 units of RNA..(pCpCpA) pyrophosphorylase. After 60 min at 37°C the reaction mixture was heated at 100°C for 1·5 min, cooled and dialysed for 48 hr against 3 changes of 6 l. of 0·2 M-NaCl and then for 24 hr against 2 changes of 6 l. of water to remove unreacted CT*P. After centrifugation to remove a small amount of insoluble material, the labeled RNA was used without further purification.

(ii) *Degradation of RNA..*pC*pC*

(a) Alkali—0·08 to 0·16 μmole of RNA..*pC*pC (about 50 μl.) was adjusted to 0·2 M-NaOH, incubated at 37°C for 20 hr, and neutralized with 1 M-HCl. A mixture of unlabeled mononucleotides was added to the digest as markers in the subsequent electrophoretic analysis.

(b) Pancreatic RNase—the reaction mixture (3·7 ml.) contained 100 μmoles of potassium phosphate buffer, pH 7·1, 18 μmoles of RNA..pCpCpA (as carrier), 7·7 μmoles of RNA..*pC*pC containing approximately 5×10^6 cts/min and 150 μg of pancreatic RNase. After 3 hr at 37°C an additional 150 μg of RNase was added and the incubation continued for 11 hr more. By this time 52% of the total *P was sensitive to PME and this did not change after further incubation.

(c) T_1 RNase—the reaction mixture (7·5 ml.) contained 400 μmoles of tris buffer, pH 7·5, 20 μmoles of EDTA, 25 μmoles of [^{14}C]RNA..*pC*pC, and 100 units of T_1 RNase. Incubation was carried out for 3 hr although the *P convertible to inorganic phosphate by *E. coli* PME reached a limit after 1 hr.

(iii) *Separation of fragments produced in the different digests*

(a) Alkali—the neutralized alkaline digests were subjected to electrophoresis on Whatman 3 MM paper in 0·05 M-ammonium formate buffer, pH 3·5, at 17 v/cm for 2 or

3 hr (Markham & Smith, 1952). Each electrophoresis strip was cut perpendicular to the direction of migration in 1 cm segments and the radioactivity of each segment was determined directly in a windowless gas-flow counter or the paper strips were eluted with water or 0·01 M-HCl and the radioactivity of the eluates measured. With the alkaline digest of [^{14}C]RNA. . .*pC*pC, the strips were counted under an aluminum shield, which eliminated $>99\%$ of the ^{14}C activity but only 70% of the ^{32}P activity. The *P activity contained within 4 mononucleotide regions (the positions were determined from the ultraviolet-absorption of the mononucleotide markers) accounted for $>97\%$ of the radioactivity on the electropherograms.

(b) Pancreatic RNase—the digest was diluted to 10 ml. with water and adsorbed on a DEAE-cellulose column ($23 \times 0{\cdot}8$ cm^2) (previously equilibrated with 0·01 M-ammonium bicarbonate, pH 8·6) and then the column was washed with 0·01 M-ammonium bicarbonate, pH 8·6. This wash fluid contained less than 2% of the *P applied to the column and was not further examined. The column was eluted first with a linear gradient of ammonium bicarbonate concentration between 0·04 M and 0·1 M, pH 8·6 (250 ml. of each solution), and then with a linear gradient between 0·1 M and 0·3 M-ammonium bicarbonate, pH 8·6 (500 ml. of each solution), at a flow rate of about 30 ml./hr. Fractions of about 5 ml. were collected, the absorbancy at 250, 260 and 280 mμ was determined and an aliquot of each fraction was dried and counted. The recovery of *P was $>95\%$ of the amount added to the column.

The fractions making up each radioactive peak were pooled, concentrated *in vacuo* to about 5 ml., lyophylized, and then kept *in vacuo* for 3 to 4 days at room temperature to remove the ammonium bicarbonate; the dried samples were dissolved in about 0·4 ml. of water.

(c) T_1 RNase—the digest was adsorbed on a column of DEAE-cellulose ($32 \times 1{\cdot}3$ cm^2) prepared as described above. The column was eluted first with a linear gradient (0·01 to 0·2 M) of ammonium bicarbonate, pH 8·6 (300 ml. of each solution), then with a second linear gradient (0·2 to 0·4 M, 500 ml. of each solution) and then with a third linear gradient (0·4 to 1 M-ammonium bicarbonate, pH 8·6, 200 ml. of each solution). Approximately 7 ml. fractions were collected and aliquots from each fraction were dried and counted with and without an aluminum shield to measure the ^{14}C- and *P-radioactivity. The fractions making up each *P-containing peak were pooled, concentrated, dried and dissolved in water as described above. The recovery of *P in the major peaks was 87%.

Each pooled fraction was subjected to paper electrophoresis at pH 3·5 and the position of *P-labeled components was determined as described above. Each *P-labeled component (most of the peaks obtained from the DEAE chromatography contained more than one *P-labeled component) was eluted from the paper strips, concentrated and chromatographed using solvent b (see below). In all cases the paper chromatogram showed a single *P-containing peak accounting for between 80 to 100% of the radioactivity on the chromatogram. The strips containing each labeled fragment were eluted with water, the eluates were concentrated to a small volume, and this solution was used for the sequence determinations.

(iv) *General methods used to identify mono-, di-, and oligonucleotides*

(a) Paper electrophoresis was performed as already described. Since authentic samples of certain di- and trinucleotides were not available the expected mobilities were calculated as described by Markham & Smith (1952).

(b) Ultraviolet absorption measurements were made in 0·01 M-HCl or 0·01 M-KOH with a Zeiss spectrophotometer. Where the ratio of absorbancy at 280 to 260 mμ for a given compound was not available from the literature it was calculated from the known absorbancies of the constituent mononucleotides (Beaven, Holiday & Johnson, 1955). In a few cases the absorbancy values were determined on eluates from paper strips; when this was done an eluate from a blank strip subjected to electrophoresis was used in the reference cuvette.

(c) Descending paper chromatography was performed at room temperature with Whatman 3 MM or Whatman no. 1 paper with the following solvents: a, isopropanol–water–ammonia (85 : 15 : 1·3 v/v); b, isobutyrate–0·5 M-ammonia (10 : 6 v/v);

c, isopropanol–acetic acid–water (60 : 30 : 10 v/v); d, isopropanol–*n*-butanol–12 N-HCl–water (114: 56 : 41 : 39 v/v). When a mixture of mononucleotides and nucleosides was to be separated, two-dimensional chromatography was used. Solvent b, which was used first, resolved all the mononucleotides except Gp and Up and all the nucleosides except G and U. The regions corresponding to these pairs were then chromatographed in the second direction using solvent a. In all cases authentic samples of the nucleotides and nucleosides were added as reference markers and visualized with a u.v. lamp. The position of the labeled compounds was determined by cutting the chromatograms into strips as already described.

(d) Enzymic degradations of a number of the isolated fragments were performed to determine the composition and sequence of nucleotides. The enzymic procedures are summarized below:

Snake venom PDE—samples in 75 μl. containing 0·1 M-glycine buffer, pH 9·2, and 0·1 M-$MgCl_2$ were incubated with 16 units of PDE at 37°C for 45 to 60 min.

E. coli *PME*—samples in 50 μl. containing 0·05 M-tris buffer, pH 8·6, and 20 units of PME were incubated at 37°C for 15 min.

B. subtilis *RNase*—samples in 50 μl. containing 0·02 M-tris buffer, pH 7·5, and 25 μg of RNase were incubated at 37°C for 30 min.

Pancreatic RNase—samples in 50 μl. containing 0·1 M-phosphate buffer, pH 7·0, were incubated with 5 μg of RNase at 37°C for 30 to 45 min.

T_1 RNase—samples in about 30 μl. containing 0·02 M-tris buffer, pH 7·5, were incubated with 8 units of RNase at 37°C for 120 min.

4. Results

(a) *Nature of the nucleotide occurring next to the CpCpA terminal sequence in amino acid-acceptor RNA chains*

Alkaline hydrolysis of RNA..*pC*pC yields a mixture of all four mononucleotides labeled with *P (Fig. 3). 50% of the *P should occur in C*p and the remaining *P

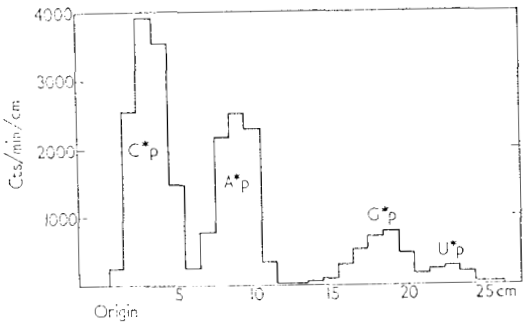

FIG. 3. Paper electrophoretic analysis of *P-labeled products produced by alkaline digestion of RNA..*pC*pC.

should be distributed amongst the nucleotides which occur next to the *pC*pC sequence (see Fig. 2). The results obtained with four preparations of RNA.. *pC*pC show that close to 50% of the *P is recovered as C*p and the remainder as A*p, G*p and U*p (Table 1). The amount of *P in C*p, in excess of 50% of the total, is a measure of the frequency of Cp as the fourth nucleotide. Since the value for C*p is so close to 50% of the total *P, little or no Cp occurs as the fourth nucleotide in acceptor RNA chains.

TABLE 1

Relative distribution of nucleotides in the fourth position of amino acid-acceptor RNA chains

Nucleotide	Distribution of *P (%)	RNA chains having indicated nucleotide in fourth position (%)
C*p	49·9 ± 1·0	< 2
A*p	34·4 ± 2·0	69
G*p	12·2 ± 2·0	24
U*p	3·6 ± 0·7	7

(b) *Identification of the fragments produced by pancreatic RNase digestion of RNA . . *pC*pC*

Figure 4 shows the results of the chromatography of the pancreatic RNase digest on DEAE-cellulose. The identification of the products of the digestion was made from the

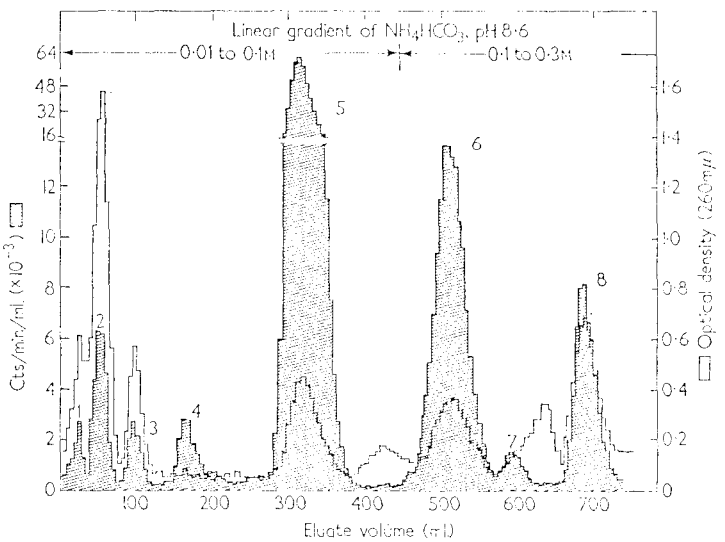

FIG. 4. Chromatographic separation of fragments produced by pancreatic RNase digestion of RNA . . *pC*pC on DEAE-cellulose.

chromatographic and electrophoretic behavior, the absorption spectra (these are summarized in Table 2) and from the products produced by chemical and enzymic degradation of the fragments as described below.

(i) rT*p (99% of peak 1) was identified on the basis of its electrophoretic mobility, and its chromatographic behavior on the DEAE column and on paper with 3 solvents. It was eluted from DEAE just before the C*p peak, which agrees with the relative positions of dpT and dpC found by Lehman (1960) with the same chromatographic system. The rT*p can be distinguished from analogues of Cp (possibly 5-methyl Cp) by its charge at pH 3·5; at this pH rT*p migrated coincident with Up as did authentic rTp and dTp. Consistent with the designation as a mononucleotide was its resistance to alkaline digestion and the

essentially complete removal of the *P by PME. Although there was some ultraviolet-absorbing material eluted from the column with the rT*p, this material was separated on paper electrophoresis. The amount of rT*p and U*p formed were about equal, yet there was considerable ultraviolet-absorbing material in the latter component. This may be explained by recognizing that the peak containing U*p (originating from the fourth nucleotide position of certain chains) also contains Up (from internal Up residues linked

TABLE 2

*Ultraviolet absorption spectra, chromatographic and electrophoretic behavior of labeled fragments produced by pancreatic RNase digestion of RNA . . .*pC*pC*

Fragment	Peak from which fragment was isolated†	Electrophoretic mobility relative to Up‡	Absorbance ratio§ (280 mμ/260 mμ)	
			pH 2	pH 12
rT*p‖	1 (>99%)	1·0	—	—
C*p¶	2 (>98%)	0·20 (0·16–0·40)	1·97 (2·0)	0·94 (0·93)
U*p††	3 (99%)	1·0	0·35 (0·32)	0·30 (0·28)
X*pC*p‡‡	4 (97%)	0·80	—	—
A*pC*p	5 (98%)	0·45 (0·43–0·50)	0·78 (0·75)	0·45 (0·41)
ApU*p	5 (1·6%)	1·1 (1·0)	—	—
G*pC*p	6 (95%)	0·72 (0·78)	1·05 (1·10)	0·75 (0·73)
ApA*pC*p	7 (95%)	0·58 (0·54)	0·53 (0·52)	0·39 (0·36)
GpA*pC*p + ApG*pC*p	8 (94%)	0·75 (0·78)	0·73 (0·75)	0·52 (0·48)

† The values in parentheses represent the amount of the component shown within the designated peak.

‡ The conditions for the electrophoresis are described in the Methods section. The values in parentheses represent previously reported or calculated values according to Markham & Smith (1952).

§ The values in parentheses are taken from the literature. Those shown for the mononucleotides are from Beaven *et al.* (1955), those for the dinucleotides and ApApCp are from Michelson (1959) and the value for the mixture of GpApCp and ApGpCp was calculated from the extinction coefficients given for the individual mononucleotides (Beaven *et al.*, 1955).

‖ rT*p chromatographed identically with authentic rTp in solvents b, c and d and with authentic Up but ahead of ψUp in solvent a. The rT*p migrated coincident with Up, ahead of ψUp and behind dTp in solvent b.

¶ C*p migrated coincident with authentic Cp on paper electrophoresis and with solvents a, b and c.

†† U*p migrated coincident with authentic Up on paper electrophoresis and with solvents a, b and c.

‡‡ X*pC*p chromatographed ahead of ψUp in solvent b and behind ψUp in solvents c and d.

to a pyrimidine). rTp, on the other hand, occurs with a frequency of about one residue per chain (Dunn, Smith & Spahr, 1960) and of these some may occur internally next to one or more purine residues thereby yielding $(Pu)_nprTp$ fragments but not free rTp. The amount of rTp occurring as the fourth nucleotide is approximately 2% of the total rTp residues.

(ii) C*p (98% of peak 2) and U*p (99% of peak 3) could not be distinguished from their authentic unlabeled counterparts by spectral, electrophoretic or chromatographic criteria. Incubation with alkali did not affect their behavior and with each compound >98% of the *P was released by PME.

(iii) X*pC*p (97% of peak 4) is a mixture of two labeled fragments. Alkaline hydrolysis yielded C*p and two labeled components 'X*p and "X*p ('X*p/"X*p = 1·0 and 'X*p+"X*p/C*p = 1·13). 'X*p and "X*p had mobilities relative to Cp of 0·54 and 0·84, respectively, in solvent b; they separated only slightly from each other in paper electrophoresis but both migrated ahead of Up. Digestion with PME yielded two labeled fragments (X*pC/*P = 1·08). There was no significant amount of ultraviolet-absorbing material associated with peak 4. This would be expected if X*pC*p contains one or more of the "unusual" purines known to occur in acceptor RNA in amounts of less than 1 residue per chain (Dunn *et al.*, 1960).

(iv) A*pC*p (98% of peak 5) yielded equal amounts of A*p and C*p after alkaline or spleen PDE hydrolysis (A*p/C*p = 0·95 and 0·97, respectively). Treatment with PME followed by alkaline hydrolysis produced A*p and *P (A*p/*P = 0·96).

(v) ApU*p (approximately 2% of peak 5) was deduced solely on the finding that about 2% of the total *P migrated as ApUp and after alkaline hydrolysis yielded only U*p. From its chromatographic behavior on DEAE it is unlikely that the fragment is GpUp or a trinucleotide terminated by Up. Insufficient material was available to make adequate spectral measurements or to determine if the fragment was AprT*p or Ap*pU*p.

(vi) G*pC*p (95% of peak 6) yielded equal amounts of G*p and C*p after alkaline or spleen PDE hydrolysis (G*p/C*p = 0·99 and 0·94, respectively). Treatment with PME followed by alkaline hydrolysis produced G*p and *P (G*p/*P = 1·01).

(vii) ApA*pC*p (95% of peak 7) yielded equal amounts of A*p and C*p after alkaline hydrolysis (A*p/C*p = 0·95). Treatment with PME followed by alkaline hydrolysis gave A*p and *P (A*p/*P = 0·97). The above structure is considered most likely based on the chromatographic, electrophoretic and spectral properties and the products produced by alkaline and enzymic hydrolysis. Excluding a tetranucleotide (because of its behavior on DEAE) the only alternative trinucleotide consistent with the specificity of pancreatic RNase is GpA*pC*p and this was identified as a component of peak 8.

(viii) GpA*pC*p and ApG*pC*p (94% of peak 8) were not resolved by paper electrophoresis at pH 3·5. Alkaline hydrolysis liberated A*p, G*p and C*p (A*p/G*p = 0·94, A*p+G*p/C*p = 1·01). Digestion with T_1 RNase yielded ApG*p, A*pC*p and C*p (C*p/A*pC*p = 1·15, C*p/ApG*p = 0·92, ApG*p/A*pC*p = 1·25). The amount of each labeled trinucleotide is obtained from the ratio of ApG*p to A*pC*p (T_1 RNase) and G*p to A*p (alkali).

(c) *Relative frequency and types of acceptor RNA chains deduced from the products of pancreatic RNase digestion*

Based on the amount and structure of the fragments produced by pancreatic RNase digestion of RNA. . *pC*pC and the reported specificity of pancreatic RNase (see Schmidt, 1955), the relative frequency of certain classes of acceptor RNA chains may be deduced (Table 3). Comparing the data in Tables 1 and 3 we can see that of the chains having A in the fourth nucleotide, (69%), approximately 94% of them are followed by a pyrimidine in the fifth position, and in only about 6% of them is A followed by a purine and then a pyrimidine. Similarly, G in the fourth position is followed predominantly by a pyrimidine and only infrequently by a purine. However, since we have accounted for only 85% of the chains containing G in the fourth

nucleotide there may yet be some fragments in which G is next to a run of purines (one such sequence, ..pGpApGpCpCpA, is suggested from the products of the T_1 RNase digest).

TABLE 3

*Terminal nucleotide sequence of acceptor RNA deduced from the fragments produced by pancreatic RNase digestion of RNA..*pC*pC*

Nucleotide fragments isolated from digest	Type of RNA chain from which fragments derived†	RNA chains having nucleotide sequence shown (%)
1. A*pC*p	..pPypApCpCpA	65·1
2. ApA*pC*p	..pPypApApCpCpA	1·1
3. GpA*pC*p	..pPypGpApCpCpA	3·1
4. G*pC*p	..pPypGpCpCpA	17·1
5. ApG*pC*p	..pPypApGpCpCpA	3·6
6. X*pC*p‡	..pPypXpCpCpA	2·8
7. U*p	..pPypUpCpCpA	2·6
8. ApU*p	..pPypApUpCpCpA	1·4
9. rT*p	..pPyprTpCpCpA	1·9
10. C*p§	..pPypCpCpA§	7·4§

† The type of RNA chain from which each fragment is derived is inferred from the specificity of pancreatic RNase and from the fact that the terminal nucleotide contains A.

‡ X*pC*p is a mixture of at least 2 types of fragments (see Table 2).

§ C*p arises from those RNA chains which contain a pyrimidine in the fourth position. Thus, the chains giving rise to U*p, rT*p and ApU*p account for 80% of those to be anticipated from the recovery of C*p.

The amount of C*p recovered in the pancreatic RNase digest should equal the frequency of pyrimidines in the fourth position, i.e. the sum of the frequencies of U*p, rT*p, and di-, tri- and oligonucleotides terminated with a Py*p. There is good agreement between the amount of C*p (7·4%) and the amount of Up in the fourth position (7%) determined by alkaline hydrolysis (Table 1), but the sum of U*p, ApU*p and rT*P account for only about 80% of the fragments of this type; possibly the remainder occurs as larger fragments not recovered from the column.

It is clear from the above results that because of the high proportion of pyrimidines in the fifth position, RNase digestion of RNA..*pC*pC is of relatively limited value for obtaining the nucleotide sequence of more than one or two residues adjacent to the pCpCpA end. Considerably more information is obtained from an analysis of the products of the T_1 RNase digestion, and this is discussed in the next section.

(d) *Products of the T_1 RNase digestion of RNA..*pC*pC*

Chromatography of the T_1 RNase digest of RNA..*pC*pC on DEAE-cellulose yields 12 peaks containing *P (Fig. 5); peaks 3 to 12 each contained more than one component. To facilitate the determination of the composition and sequence of nucleotides in each fragment, the RNA..*pC*pC contained ^{14}C in each of the bases except the two terminal cytidylate residues. The unshaded profile, which represents fragments derived from internal positions of the RNA chains, is interchangeable with the optical density.

Before presenting the analytical data for the different fragments it is important to outline the approaches used, together with some of the uncertainties which exist. T_1 RNase digestion of RNA..*pC*pC produces several different types of *P-labeled fragments (see Fig. 2). The first type is terminated by G*p and arises from all chains in which the nucleotide next to the C*pC end contains G. The length of these fragments is determined by the spacing between this G and the next G residue in the

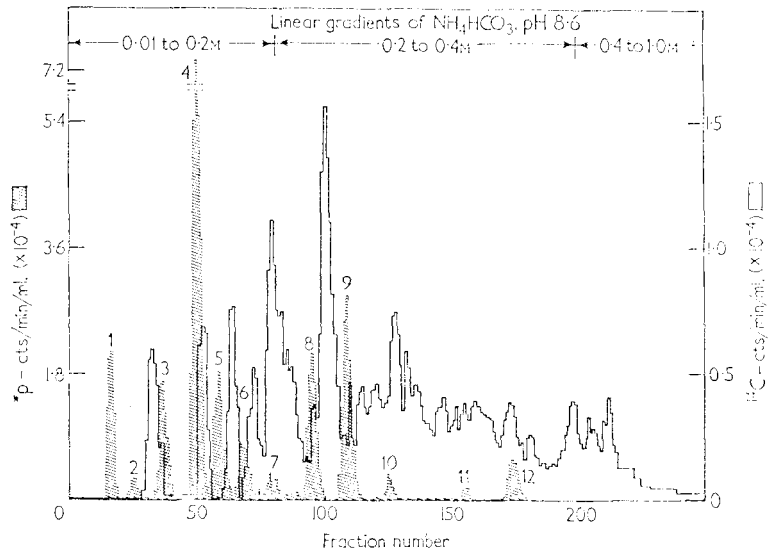

FIG. 5. Chromatographic separation of fragments produced by T_1 RNase digestion of RNA..*pC*pC on DEAE-cellulose.

chain. All chains which contain G in the fourth position also give rise to a common fragment, C*pC; the amount of C*pC formed is, therefore, an independent measure of the number of chains with G in the fourth nucleotide. The other type of labeled fragment is terminated by ..*pC*pC and is produced from all chains having A or U in the fourth nucleotide; the length of these fragments is a measure of the position of the first G in the chain.

A schematic illustration of the approach used in analysing fragments of the types mentioned above is summarized in Fig. 6.

(i) *Fragments terminated by *pC*pC (Fig. 6(a))*

(a) The *P-containing products produced by alkaline digestion (i) define the nature of the nucleotide adjacent to the C*pC unit (i.e. the fourth nucleotide of the chain from which the oligonucleotide is derived).

(b) Alkaline, followed by PME digestion (ii), yields inorganic phosphate and a mixture of [^{14}C]nucleosides; the types and the proportion of nucleosides give the composition of the fragment.

(c) Degradation of the fragment with venom PDE (iii) yields a mixture of 5′-mononucleotides and a nucleoside corresponding to the 5′-hydroxy terminal nucleotide of the fragment.

(d) Pancreatic RNase digestion (iv) serves to split the relatively long fragment into shorter ones. With the fragment shown in Fig. 6(a) a *P-containing ApCp fragment (●A*pC*p) and ApCp with no *P (●Ap●Cp) are produced; the analysis of this mixture of dinucleotides serves to confirm the sequence deduced by the other methods.

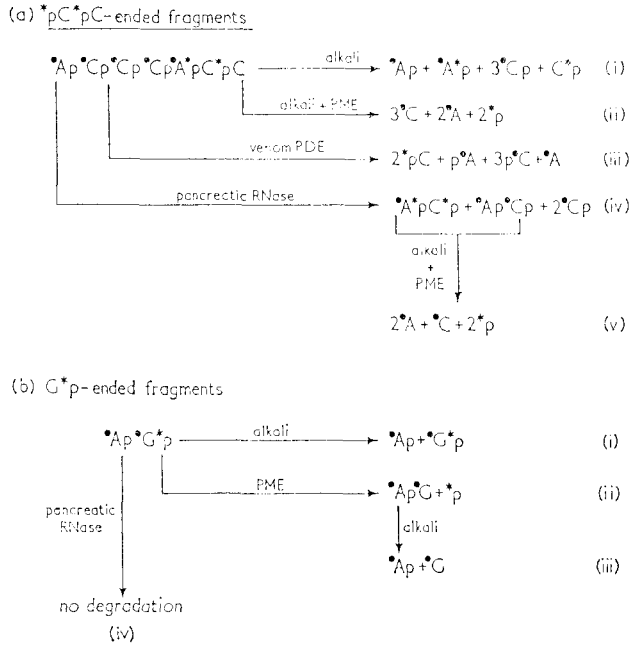

FIG. 6. General plan for nucleotide composition sequence analysis of: (a)..*pC*pC and (b) G*p-ended fragments produced by T_1 RNase digestion of RNA..*pC*pC.

(ii) *Fragments terminated by G*p (Fig. 6(b))*

(a) The finding of G*p after alkaline digestion (i) defines this type of fragment as one which originates from a chain having G adjacent to the CpC end.

(b) Further proof that it is a G*p-ended fragment comes from the liberation of ●G after PME action (ii), followed by alkaline digestion (iii). In the example shown in Fig. 6(b) the ratio ●Ap/●G and the lack of any other labeled nucleotides define the composition and the length of the fragment.

(c) The failure to observe degradation with pancreatic RNase (iv) eliminates the existence of a pyrimidine in the fragment and further supports the conclusion that the Ap is linked to the Gp.

(iii) *Uncertainties in characterization of fragments*

One of the difficulties encountered in determining the nucleotide sequences of the isolated fragments was contamination of the *P-labeled fragments (which also contain ^{14}C) with fragments containing only ^{14}C. To appreciate this problem consider first a fragment of the type shown in Fig. 6(a); this type does not arise from any portion of the RNA chains except from the labeled end. Therefore the concentration of these in the digest is small (1 per chain) but they are identifiable by their *P content. Since the specific activity of the *P is approximately 10 times higher than the ^{14}C specific activity of each nucleotide, a fragment of the type shown in Fig. 6(a) contains

approximately 4 times more *P than ^{14}C. The shorter the fragment the greater the ratio $*P/^{14}C$. Therefore contamination of such a fragment with another fragment containing only ^{14}C is extremely difficult to detect during the isolation and purification procedure. The consequences of such contamination become apparent in analysing the composition and sequence of the nucleotides in the fragment. For example, after cleavage of the C*pC-ended fragment (Fig. 6(a), ii) to its constituent mononucleotides and removal of the phosphate groups, contamination by, for example, $^{\bullet}$Up$^{\bullet}$Gp would yield $^{\bullet}$U and $^{\bullet}$G which do not occur in the *P-labeled fragment.

Fragments terminated by Gp present a somewhat different situation. Such fragments comprise by far the predominant type since they arise from any region of the RNA chains; thus the fragment $^{\bullet}$Ap$^{\bullet}$Gp may occur many times throughout all the chains and these of course will be isolated along with the $^{\bullet}$Ap$^{\bullet}$G*p fragment linked to the CpC end. In this case, depending upon the amount of $^{\bullet}$Ap$^{\bullet}$Gp produced, the ^{14}C content may be many times higher than the *P level. While the $*P/^{14}C$ ratio is a criterion of purity during the isolation of the fragment, somewhat different uncertainties arise.

The first of these is the determination of the length of the fragment. Consider the fragment $^{\bullet}$Ap$^{\bullet}$G*p which, for example, may be mixed with $^{\bullet}$Ap$^{\bullet}$Gp in the ratio of 1 to 10. Whereas the amount of *P in a fragment terminated with *pC*pC is a measure of the amount of the fragment, the amount of *P in the G*p-ended fragment is not; in this example there is 10 times more of the fragment than is indicated by the *P. However, since G is the only base which cannot occur more than *once* in such fragments, removal of the terminal phosphate permits the use of $^{\bullet}$G to establish the quantitative relationships with the other nucleotides in the fragment.

The second uncertainty stems from a possible failure to separate completely two different fragments each terminated with Gp. Consider, for example, a mixture of 20 parts $^{\bullet}$Cp$^{\bullet}$Ap$^{\bullet}$Gp and 1 part $^{\bullet}$Cp$^{\bullet}$Cp$^{\bullet}$Cp$^{\bullet}$G*p. After removal of the terminal phosphate, such a mixture would yield on alkaline hydrolysis, $^{\bullet}$Cp, $^{\bullet}$Ap and $^{\bullet}$G in the approximate ratio of 1 : 0·87 : 0·91. It is quite clear that in this instance it would be erroneous to conclude that the sequence CpApGp occurs next to the labeled CpC end group.

(e) *Identification of the fragments produced by T_1 RNase digestion of RNA..*pC*pC*

Table 4 summarizes the electrophoretic and chromatographic properties of the *P-labeled fragments isolated from the digest. The procedures and the data upon which the identification of these fragments is based are as follows in Table 4 overleaf:

TABLE 4

*Electrophoretic and chromatographic behavior of *P-labeled fragments produced by T_1 RNase digestion of RNA..*pC*pC*

	Fragment	Peak of origin	Electrophoretic mobility†	Chromatographic mobility in solvent b
			(cm)	(migration relative to Cp)
1.	C*pC	1	0	—
2.	•A*pC*pC	3	1–2	1·1 –1·2
3.	•Cp•A*pC*pC	4	5–6	0·95–1·0
4.	•Cp•U*pC*pC	5	10–11	0·60–0·70
5.	•Up•A*pC*pC	5	13–14	0·70–0·80
6.	•Cp•Cp•U*pC*pC	6	9–10	0·35–0·45
7.	•Cp•Cp•A*pC*pC	6	7	0·70–0·80
8.	•Ap•G*p	7	25–26	0·70–0·80
9.	•Cp(•Up•C)p•A*pC*pC‡	8	14–15	0·50–0·60
10.	•Cp(•Up•Cp•C)p•A*pC*pC‡	9	15–16	0·35–0·45
11.	•Ap•Cp•Cp•Cp•A*pC*pC	9	11–12	0·70–0·80
12.	•Up(•Cp•Cp•U)p•G*p‡	11	32	0·15–0·20

† The electrophoretic mobilities are not necessarily the true mobilities of the fragment shown, since they were measured with the fractions recovered from the DEAE column which contained variable amounts of residual ammonium bicarbonate. They are useful, however, in showing the separation and relative mobilities of those fragments which were eluted from the DEAE column in the same peak.

‡ The nucleotide orders within the parentheses are arbitrary and have not yet been experimentally determined.

(1) C*pC on alkaline hydrolysis yielded C*p (100% of the total *P) and a small amount of •C (•C/C*p = 0·06). Treatment with venom PDE gave *pC (100% of total *P) and •C/*pC = 0·11.

(2) •A*pC*pC on alkaline hydrolysis gave A*p/C*p = 0·98. Alkaline hydrolysis followed by PME digestion produced •A/*P = 0·66. Digestion with *B. subtilis* RNase yielded A*p/C*pC = 0·77 and treatment with venom PDE gave •A/*pC = 0·52.

(3) •Cp•A*pC*pC after hydrolysis with alkali gave A*p/C*p = 0·98 and sequential degradation with alkali and PME gave •A/*P = 0·41 and •C/*P = 0·49. Treatment with pancreatic RNase yielded •Cp/A*pC*p = 0·92 and a trace of •C (•C/A*pC*p = 0·07). Degradation with venom PDE gave •C/*pC = 0·41 and p•A/*pC = 0·47.

(4) •Cp•U*pC*pC on alkaline hydrolysis yielded U*p/C*p = 1·05. Following alkaline digestion, treatment with PME gave •C/*P = 1·05, •U/*P = 0·68. Because the values of •C/*P and •U/*P were higher than expected for the structure shown, the fragment was rechromatographed in solvent b. Treatment of the recovered material with pancreatic RNase yielded U*p/C*p = 1·0; Up and Cp isolated from this digest treated with PME gave •U/*P = 0·90 and •C/*P = 1·08.

(5) •Up•A*pC*pC when hydrolysed with alkali produced A*p/C*p = 0·95. Alkaline hydrolysis followed by PME digestion gave •A/*P = 0·58, •U/*P = 0·53, and •C/*P = 0·19. Treatment with pancreatic RNase gave •Up/A*pC*p = 0·92 and •C/A*pC*p = 0·06.

(6) •Cp•Cp•U*pC*pC upon alkaline hydrolysis gave U*p/C*p = 0·90 and sequential alkaline and PME hydrolysis yielded •U/*P = 0·53, •C/*P = 1·3. Pancreatic RNase digestion produced U*p/C*p = 1·1; Up and Cp isolated from RNase digest and treated with PME produced •U/*P = 1·2, •C/*P = 2·5, •C/•U = 2·1.

(7) •Cp•Cp•A*pC*pC produced on alkaline hydrolysis A*p/C*p = 0·92 and PME treatment of the alkaline digest gave •A/*P = 0·48, •C/*P = 1·17. Pancreatic RNase digestion yielded •Cp/A*pC*p = 2·0.

(8) •Ap•G*p on hydrolysis gave G*p (>95% of the total *P), •Ap/•Gp = 1·17. Digestion with PME yielded •Ap•G which on alkaline hydrolysis yielded •Ap/•G = 0·88. The fragment was unchanged after pancreatic RNase treatment.

(9) •Cp(•Up•C)p•A*pC*pC on hydrolysis with alkali produced A*p/C*p = 1·0. PME digestion following alkaline hydrolysis produced •A/*P = 0·48, •U/*P = 0·49, •C/*P = 1·1, and •G/*P = 0·08. Digestion with venom PDE gave •C/*pC = 0·49, p•U/*pC = 0·57, p•A/*pC = 0·55 and p•C/p*C = 0·85. Digestion with pancreatic RNase produced •Cp/A*pC*p = 2·4 and •Up/A*pC*p = 0·92.

(10) •Cp(•Up•Cp•C)p•A*pC*pC on alkaline hydrolysis yielded A*p/C*p = 0·97. Alkaline hydrolysis followed by PME gave •A/*P = 0·46, •U/*P = 0·46, •C/*P = 1·53. Pancreatic RNase digestion produced •Cp/A*pC*p = 3·0, •Up/A*pC*p = 1·0. Digestion with venom PDE yielded •C/*pC = 0·62, p•U/*pC = 0·74, p•A/*pC = 0·53 and p•C/*pC = 1·0.

(11) •Ap•Cp•Cp•Cp•A*pC*pC after alkaline hydrolysis gave A*p/C*p = 0·95, and sequential hydrolysis with alkali and PME gave •A/*P = 1·1, •C/*P = 1·3 and •G/*P = 0·07. Pancreatic RNase digestion gave •Cp/A*pC*p = 2·04; the isolated mixture of •Ap•Cp and •A*pC*p after hydrolysis with alkali and PME produced •A/*P = 0·99, •C/*P = 0·52, •A/•C = 1·90. Treatment with venom PDE gave •A/*pC = 0·6, p•A/*pC = 0·64, p•C/*pC = 1·7.

(12) •Up(•Cp•Cp•U)p•G*p after alkaline hydrolysis gave G*p (>95% of the total *P). Treatment of the fragment with PME removed the *P (>95%) producing •Up(•Cp•Cp•U)p•G. Exposure of the dephosphorylated fragment to venom PDE gave p•G/•U = 0·83, p•C/•U = 2·04 and p•U/•U = 1·0.

In most cases the results of the chemical and enzymic degradation of the fragments agree reasonably well with the predicted values for the sequences shown. With fragment 9 digestion with venom PDE gave a somewhat higher ratio p•C/*pC than is predicted (0·5) from the postulated structure and this was also evident in the elevated ratio of •Cp/A*pC*p after pancreatic RNase digestion. It is unlikely that fragment 9 has an additional •Cp residue in the chain since this would make it equivalent to fragment 10, the composition of which is more firmly established.

None of the derivatives of Up has been found in the fragments so far identified. In every case where Up occurred, it could be distinguished from ψUp or rTp by comparing the isolated labeled nucleoside, •U, with authentic ψU and rT in the two-dimensional chromatographic system (solvent b followed by solvent a). No attempt was made in these analyses to detect the known analogues of the purines (Dunn *et al.*, 1960) in the isolated fragments.

(f) *Relative frequency and types of acceptor RNA chains deduced from the products of T_1 RNase digestion*

Table 5 shows the different types and relative proportions of acceptor RNA chains, which correspond to the structure of the identified fragments. Implicit in these designations is the assumption that each fragment resulted from cleavages between guanosine-3′-phosphate and the 5′-hydroxyl group of the adjacent nucleotide (Sato-Asano, 1959).

Approximately 80% of the acceptor RNA chains in the mixed population can be accounted for by the sequences shown. Assuming that the frequency of each of the classes of RNA chains is the same in the RNA preparations isolated from *E. coli* B and from *E. coli* strains W3362 and W3687† we may compare the information obtained from the pancreatic and T_1 RNase digestions. Of the acceptor RNA chains having A

† The pancreatic RNase digestion and analysis was carried out with RNA..*pC*pC originally isolated from *E. coli* B while the RNA used to prepare the RNA..*pC*pC for T_1 RNase digestion was isolated from *E. coli* strain W.

in the fourth position, (69%), >90% of these can be accounted for by the sequences shown. Moreover, as suggested by the products of the pancreatic RNase digestion, all but one of these (sequence 2) are characterized by having a pyrimidine in the fifth nucleotide. Of the RNA chains corresponding to fragment 2 some may have a pyrimidine following the G in the fifth position; these would correspond to the chains giving rise to fragment 3 recovered from the pancreatic RNase digest (Table 3). Although the sequence ..PypApApCpCpA can be inferred from the existence of ApA*pC*p in the pancreatic RNase digest, no corresponding fragment, i.e. ..PypA*pC*pC, has yet

TABLE 5

*Terminal nucleotide sequences of acceptor RNA deduced from the fragments produced by T_1 RNase digestion of RNA..*pC*pC*

	Nucleotide fragments isolated from digest	Type of RNA chain from which fragment derived	RNA chains having nucleotide sequence shown (%)
1.	CpC	..GpCpCpA†	24†
2.	ApCpC	..GpApCpCpA	8
3.	CpApCpC	..GpCpApCpCpA	25
4.	CpUpCpC	..GpCpUpCpCpA	6
5.	UpApCpC	..GpUpApCpCpA	3
6	CpCpUpCpC	..GpCpCpUpCpCpA	2
7.	CpCpApCpC	..GpCpCpApCpCpA	4
8.	ApGp	..GpApGpCpCpA	4
9.	Cp(UpC)pApCpC‡	..GpCp(UpC)pApCpCpA	8
10.	Cp(UpCpC)pApCpC‡	..GpCp(UpCpC)pApCpCpA	7
11.	ApCpCpCpApCpC	..GpApCpCpCpApCpCpA	9
12.	Up(CpCpU)pGp‡	..GpUp(CpCpU)pGpCpCpA	5

† This class of RNA chains includes all chains having Gp as the fourth nucleotide. Less than half of these is accounted for by the sum of fragments 8 and 12.

‡ The nucleotide orders within the parentheses are arbitrary and have not yet been experimentally determined.

been found in the T_1 RNase digest. Of the RNA chains having U in the fourth position almost all have been accounted for, although fragments derived from chains with rT in the fourth nucleotide or those with the terminal sequence ..ApUpCpCpA have not yet been found. There is good agreement between the amount of C*pC produced in the T_1 RNase digest with the amount of RNA chains having G in the fourth nucleotide (see Table 1). However, only about 35 to 40% of the predicted amount of Gp-ended fragments have been accounted for thus far; one of these has a pyrimidine in the fifth nucleotide and the other has A in the fifth position.

5. Discussion

The specificity of each amino acid-acceptor RNA chain very likely results from differences in the nucleotide sequence within each chain. One may ask if the differences in nucleotide sequence are confined to a particular segment of each chain or if they are distributed throughout the chain. The present study shows that there are striking differences in the nucleotide sequences adjacent to the common terminal trinucleotide, ..pCpCpA. The heterogeneity of nucleotide sequences in this region is already apparent in the fourth nucleotide in the chain, i.e. the nucleotide adjacent to the pCpCpA end

group. A is the predominant representative in this position; G and smaller amounts of U and rT also occur but C is not found in the fourth position to a significant extent. No *U was detected next to the pCpCpA end group, although two unidentified nucleotides, very likely purines, appear in the fourth position of about 3% of the RNA chains. Canellakis & Herbert (1961), using rat liver-acceptor RNA, have reported that the nucleotide adjacent to the pCpCpA end group is almost exclusively A.

The frequency of the different nucleotides in the fourth position deviates significantly from that predicted from the overall base composition of the acceptor RNA (Ofengand *et al.*, 1961). For example, since G and C occur to the greatest extent and in almost equal amounts they might, on a random basis, be expected to occur with equal frequency in the fourth position. Similarly, A should have occurred less often than G or C and close to the frequency of U. Moreover, in the population of acceptor RNA chains the ratio of purines to pyrimidines is very nearly 1 (Ofengand *et al.*, 1961; Dunn *et al.*, 1960) yet the ratio of the frequencies of purine to pyrimidine in the fourth position is nearly 10. One would also predict, on the basis of random sequences, to find almost equal amounts of ..PupApCpCpA and PypApCpCpA but we find over 15 times more of the ..PypAp.. sequence than of ..PupAp...

In the T_1 RNase digest of RNA.. *pC*pC almost all the chains having A and U in the fourth nucleotide are accounted for by the fragments having *pC*pC end groups. On the other hand, only 35 to 40% of the fragments terminated by ..G*p have been isolated although, based on the recovery of C*pC, the digestion was complete. The Gp-ended fragments are more difficult to isolate because they are not easily separated from the large number of different Gp-ended fragments originating from internal positions of the chains. G*p, which would arise from chains of the type RNA..pGpG*pC*pC, has not been observed suggesting that little or none of this sequence occurs.

Although the present study shows that the chains vary in the nucleotide sequences occurring next to the CpCpA end group, we do not know whether this region contributes structural information to distinguish the different chains. In this connection we have determined that there are 20 to 25 different *P-labeled fragments in the T_1 RNase digest, implying at least that many different chains. One of the sequences, RNA..pGpCpApCpCpA represents about 25% of the chains in the population, and there is no indication that any one amino acid is bound to that extent. In the next paper (Berg *et al.*, 1962) evidence is presented which suggests that leucine is bound in part to chains of the type just mentioned, yet the leucine-specific acceptor chains do not account for more than about 40% of this class. We may suppose therefore that in this case the three nucleotides following the CpCpA sequence may be common to chains with different amino acid specificity. Since there are indications that differences exist between amino acid-specific acceptor RNA chains from different species (Berg *et al.*, 1961; Benzer & Weisblum, 1961; Allen, Glassman, Cordes & Schweet, 1960), it would be of interest to compare the types of terminal nucleotide sequences which occur in such RNA preparations.

To evaluate the results reported here it is important to consider errors which could influence the amount and position of the *P introduced into the RNA and thereby affect the types and distribution of the labeled fragments found in the digests. These errors are:

(i) Addition of CM*P onto the ends of polyribonucleotides other than amino acid-acceptor RNA or onto oligo- or polynucleotides derived from partially degraded

acceptor RNA; both occurrences would introduce the label adjacent to nucleotides not normally occurring at the end of the acceptor RNA chains. In considering these possibilities it should be emphasized that such extraneous CM*P acceptors would have to possess unesterified terminal 3′-hydroxyl groups.† Experiments bearing on this question show that intact or partially degraded ribosomal RNA is not pyrophosphorolysed nor does either function as acceptors of CMP or AMP from the respective nucleoside triphosphates (Preiss *et al.*, 1961; Furth *et al.*, 1961). Poly A, poly AGUC and poly AU as well as RNA from tobacco or turnip yellow mosaic viruses are not active as nucleotide acceptors or detectably pyrophosphorolysed. Earlier studies with venom PDE showed that, whereas removal of all or a part of the pCpCpA segment from intact acceptor RNA was essential for incorporating CMP or AMP into the RNA, removal of additional nucleotides from the chains eliminated the CMP-acceptor activity (Preiss *et al.*, 1961). Internal cleavages in acceptor RNA chains produced by the PDE action of an RNase from *E. coli* (Spahr & Illingworth, 1961) lead to simultaneous loss of amino acid-acceptor activity and the ability to be pyrophosphorolysed (Berg, unpublished experiments). These experiments suggest that RNA... (pCpCpA) pyrophosphorylase requires an intact RNA chain for pyrophosphorolysis; to incorporate CMP it requires in addition the unesterified 3′-hydroxyl group of the nucleotide normally adjacent to the CpCpA end group or of a cytidylate residue linked to that nucleotide.

(ii) Failure to remove quantitatively the terminal pCpCpA segment and/or to label subsequently each chain with two CM*P residues. The extent of the error introduced by either of the above depends to a great extent on whether the unlabeled or incompletely labeled chains are randomly or non-randomly distributed amongst the population. If they are randomly distributed then the yield of labeled fragments is decreased but the types of fragments and their relative frequencies in the population will be unaffected. If, however, only certain types of chains fail to be labeled then our estimate of the types and frequencies of different sequences will be in error and there may even be an entire class of sequences which would be missed.

The extent of pyrophosphorolysis of the RNA can be evaluated by examining the amount of CMP which can be added back. This has averaged 2 CMP residues per average chain length of 89 nucleotides which agrees closely with estimates of the molecular weight of the acceptor RNA (Ofengand *et al.*, 1961). Moreover, each chain which can accept any CMP appears capable of accepting 2 CMP residues since the ratio of CMP to AMP which can be added back after pyrophosphorolysis is very close to 2. If there were a significant class of RNA chains from which only AMP and 1 CMP residue had been removed, then the ratio of CMP to AMP incorporation would be less than 2 (Preiss *et al.*, 1961).

(iii) Failure of pancreatic and T_1 RNase to degrade the RNA chains completely or degradation of the chains at positions other than those predicted from the reported enzyme specificity. In this regard none of the fragments identified thus far, from either the pancreatic or T_1 RNase digest, contained a linkage susceptible to the nuclease used to produce the digest. In the pancreatic RNase digest only C*p, U*p, rT*p and di- and trinucleotides terminated with a Py*p group (the total accounting

† To have produced such fragments by inadvertent enzymatic cleavages would require an RNA diesterase with a specificity for cleaving between the 3′-hydroxyl group and the phosphate, i.e. liberating 5′-phosphomonoesters. Only two such degradative enzymes acting on RNA are known, snake venom PDE (Schmidt, 1955) and an RNA phosphodiesterase from *Azotobacter vinlandii* (Stevens & Hilmoe, 1960).

for better than 95% of the total *P in the RNA) were observed. Similarly, of the identified fragments produced in the T_1 RNase digestion, which account for about 80% of the *P, all were terminated either by the expected ..*pC*pC or G*p group. Moreover, no *P-containing purine mononucleotides were detected in the pancreatic RNase digest; there were also no *P-containing mononucleotides found in the T_1 RNase digest, although G*p was a conceivable product (see earlier part of discussion). Such labeled mononucleotides would have been indicative of abnormal cleavages. Thus, we have no reason at present to suspect either cleavages at points other than those predicted from the reported specificity of the two ribonucleases or that susceptible linkages were not hydrolysed. In any case, the major error introduced by an abnormal cleavage is in the designation of the innermost nucleotide shown in the sequences presented in Tables 2 and 5.

This investigation was supported by grant funds from the National Institutes of Health of the U.S. Public Health Service.

We wish to express our gratitude to Dr. T. Breitman for the ψUp, to Professor F. Egami, Tokyo, Japan, for providing the takadiastase (Sankyo), to Dr. E. Lederberg for providing the purineless and pyrimidineless mutants of *E. coli*, and Dr. S. Nishimura, Tokyo, Japan, for a gift of the *B. subtilis* RNase. We should also like to acknowledge the assistance of Mr. John Milligan in carrying out the electrophoretic and chromatographic analyses of some of the fragments.

REFERENCES

Allen, E. H., Glassman, E., Cordes, E. & Schweet, R. S. (1960). *J. Biol. Chem.* **235**, 1068.
Allen, E. H., Glassman, E. & Schweet, R. S. (1960). *J. Biol. Chem.* **235**, 1061.
Beaven, G. H., Holiday, E. R. & Johnson, E. A. (1955). In *The Nucleic Acids*, ed. by E. Chargaff & J. N. Davidson, vol. I, p. 493. New York: Academic Press.
Benzer, S. & Weisblum, B. (1961). *Proc. Nat. Acad. Sci., Wash.* **47**, 1149.
Berg, P. (1961). *Ann. Rev. Biochem.* **30**, 293.
Berg, P., Bergmann, F. H., Ofengand, E. J. & Dieckmann, M. (1961). *J. Biol. Chem.* **236**, 1726.
Berg, P., Lagerkvist, U. & Dieckmann, M. (1962). *J. Mol. Biol.* **5**, 159.
Berg, P. & Ofengand, E. J. (1958). *Proc. Nat. Acad. Sci., Wash.* **44**, 78.
Canellakis, E. S. & Herbert, E. (1960). *Proc. Nat. Acad. Sci., Wash.* **46**, 170.
Canellakis, E. S. & Herbert, E. (1961). *Biochim. biophys. Acta*, **47**, 78.
Dunn, D. B., Smith, J. D. & Spahr, P. F. (1960). *J. Mol. Biol.* **2**, 113.
Furth, J. J., Hurwitz, J., Krug, R. & Alexander, M. (1961). *J. Biol. Chem.* **236**, 3317.
Garen, A. & Levinthal, C. (1960). *Biochim. biophys. Acta*, **38**, 470.
Hecht, L. I., Stephenson, M. L. & Zamecnik, P. C. (1959). *Proc. Nat. Acad. Sci., Wash.* **45**, 505.
Hecht, L. I., Zamecnik, P. C., Stephenson, M. L. & Scott, J. F. (1958). *J. Biol. Chem* **233**, 954.
Hoagland, M. B. (1960). In *The Nucleic Acids*, ed. by E. Chargaff & J. N. Davidson, vol. III, p. 349. New York: Academic Press.
Holley, R. W., Apgar, J., Merrill, S. H. & Zubkoff, P. L. (1961). *J. Amer. Chem. Soc.* **83**, 4861.
Hurwitz, J. (1959). *J. Biol. Chem.* **234**, 2351.
Josse, J., Kaiser, A. D. & Kornberg, A. (1961). *J. Biol. Chem.* **236**, 864.
Koerner, J. F. & Sinsheimer, R. L. (1957). *J. Biol. Chem.* **228**, 1049.
Lehman, I. R. (1960). *J. Biol. Chem.* **235**, 1479.
Lehman, I. R., Bessman, M. J., Simms, E. S. & Kornberg, A. (1958). *J. Biol. Chem.* **233**, 163.
Lipmann, F., Hülsmann, W. C., Hartmann, G., Boman, H. G. & Acs, G. (1959). *J. Cell. Comp. Physiol.*, **54**, Suppl. 1, 75.

Markham, R. & Smith, J. D. (1952). *Biochem. J.* **52**, 558.
Michelson, A. M. (1959). *J. Chem. Soc.*, 3655.
Nishimura, S. (1960). *Biochim. biophys. Acta*, **45**, 15.
Ofengand, E. J. (1959). Ph.D. thesis, Washington University, St. Louis, Missouri.
Ofengand, E. J., Dieckmann, M. & Berg, P. (1961). *J. Biol. Chem.* **236**, 1741.
Preiss, J., Berg, P., Ofengand, E. J., Bergmann, F. H. & Dieckmann, M. (1959). *Proc. Nat. Acad. Sci., Wash.* **45**, 319.
Preiss, J., Dieckmann, M. & Berg, P. (1961). *J. Biol. Chem.* **236**, 1748.
Sato-Asano, K. (1959). *J. Biochem.* **46**, 31.
Schmidt, G. (1955). In *The Nucleic Acids*, ed. by E. Chargaff & J. N. Davidson, vol. I, p. 555. New York: Academic Press.
Singer, M. F. & Cantoni, G. L. (1960). *Biochim. biophys. Acta*, **39**, 182.
Spahr, P. F. & Illingworth, B. R. (1961). *J. Biol. Chem.* **236**, 823.
Stevens, A. & Hilmoe, R. J. (1960). *J. Biol. Chem.* **235**, 3023.
Takahashi, K. (1961). *J. Biochem.* **49**, 1.
Zachau, H. G., Acs, G. & Lipmann, F. (1958). *Proc. Nat. Acad. Sci., Wash.* **44**, 885.
Zillig, W., Schaetschabel, D. & Krone, W. (1960). *Z. physiol. Chem.* **318**, 100.